Variational Quantum Algorithms (VQAs)

Topics to cover:

VQAs for ground state preparation

- Variational Quantum Eigensolver:
  - find ground state energies of Hamiltonian H by minimizing $E(\theta)$
  - example: Kitaev honeycomb spin model $H_K = J \sum_{\langle ij \rangle \alpha} S_i^\alpha S_j^\alpha$
  - gradient-based vs. non-gradient based optimizers
  - how to compute gradients using parameter-shift rules
  - Quantum natural gradient (QNG) & interpretation
  - QNG & relation to VQITE
- VQITE based on MacLachlan principle (derivation)
- AVQITE: adaptive ansatz generation
- Shot distribution: Kuczor approach (?)
- Brief detour: preparation of excited & thermal states

VQAs for real-time dynamics

- MacLachlan dynamics (VQDS): $M \dot{\theta} = V$

- circuit example for TFIM
- global phase as a parameter
- fixed ansatz versus adaptive ansatz generation

• Variational fast forwarding (VFF)

• Circuit compression
  - integrable models
  - VTC algorithm

What is a variational quantum algorithm?

Key idea: represent a quantum state of interest via a parametrized quantum circuit (= parametrized unitary operator):

reference state, e.g., $|0\rangle$.

$$|\psi(\vec{\theta})\rangle = U(\vec{\theta})\,|\psi_0\rangle$$

[$\vec{\theta}$ is a classical representation of the quantum state $|\psi(\vec{\theta})\rangle$.]

Then, determine parameters $\vec{\theta}$ by classically optimizing an objective cost function $C(\vec{\theta})$, which can be computed by preparing $|\psi(\vec{\theta})\rangle$ on a QC and measuring expectation values. Typically choose $U(\vec{\theta})$ to consist of single & two-qubit gates.

VQAs are ideally tailored to NISQ conditions as we restrict $U(\vec{\theta})$ to circuits that can be efficiently implemented on hardware.

Caveat 1: only certain states can be reached with shallow circuits, but circuit complexity $\neq$ entanglement (unlike in classical TN approaches)

Caveat 2: trade circuit depth for number of measurements (often this is a bottleneck in practice).

Examples:

Hamiltonian (or general hermitian operator = observable)

① $\cdot\ C(\vec{\theta}) = \langle\psi(\vec{\theta})|\, H \,|\psi(\vec{\theta})\rangle \equiv E(\theta)$

$\rightarrow$ goal is to find GS energy.

$$E_{GS} \leq \min_{\vec{\theta}} E(\vec{\theta}) \quad \text{[variational principle]}$$

classical minimization over noisy cost function!

This is VQE.

nonlocal even if H is local

② $\cdot\ C(\vec{\theta}) = \langle\psi_\theta|\, H^2 \,|\psi_\theta\rangle - \langle\psi_\theta|H|\psi_\theta\rangle^2$

minimize the energy variance $\Rightarrow$ prepare any eigenstate, also (highly) excited ones.

This is VQE-X.

(3) $\cdot\ C(\vec{\theta}) = \langle\psi_\theta|\overbrace{(H-\lambda)^2}^{=H^2-2\lambda H+\lambda^2}|\psi_\theta\rangle =$

$$= \langle\psi_\theta|H^2|\psi_\theta\rangle - \langle\psi_\theta|H|\psi_\theta\rangle^2$$

$$+ \langle\psi_\theta|(H-\lambda)|\psi_\theta\rangle^2$$

Minimize sum of energy variance and energy difference to specified target λ (can be generalized

This is the folded spectrum method.

(4) $\cdot\ C(\vec{\theta}) = 1 - |\langle\psi_{target}|\psi_\theta\rangle|^2$

Minimize infidelity with a target state.

$\Rightarrow C(\vec{\theta}) = 1 - |\langle\psi_0|U^\dagger_{target}\, U_\theta|\psi_0\rangle|^2 = 1 - p_{\psi_0}$

$\Rightarrow$ measure probability to return to $|\psi_0\rangle$

$\hat{=}$ measure global projector $P_G = \mathbb{1} - |\psi_0\rangle\langle\psi_0|$:

$$C(\vec{\theta}) = \langle \psi_0 | \mathcal{U}_\theta^\dagger \mathcal{U}_T \left(\mathbb{1} - |\psi_0\rangle\langle\psi_0| \right) | \mathcal{U}_T^\dagger \mathcal{U}_\theta |\psi_0\rangle$$
$$= 1 - |\langle \psi_0 | \mathcal{U}_T^\dagger \mathcal{U}_\theta | \psi_0 \rangle|^2 = 1 - p_{\psi_0}$$

Global observables often difficult to optimize for large systems (one out of exponentially many states, barren plateaus)

Alternative: local cost function that exhibits the same minimum: $\underset{\vec{\theta}}{\operatorname{argmin}}\, C_G(\vec{\theta}) = \underset{\vec{\theta}}{\operatorname{argmin}}\, C_L(\vec{\theta})$

$$P_L = \mathbb{1} - \frac{1}{N} \sum_{j=1}^{N} |\psi_0\rangle_j\, {}_j\langle\psi_0| \otimes \mathbb{1}_{\bar{k}}$$

identity matrix for all qubits except k_j

Example: $|\psi_0\rangle = |0000\rangle$

$$\mathcal{U}_T^\dagger \mathcal{U}_\theta |\psi_0\rangle = \sqrt{p_0}\,|0000\rangle + \sqrt{p_2}\,|0010\rangle + \sqrt{p_5}\,|0101\rangle + \sqrt{p_8}\,|1000\rangle + \sqrt{p_{14}}\,|1110\rangle + \sqrt{p_{15}}\,|1111\rangle$$

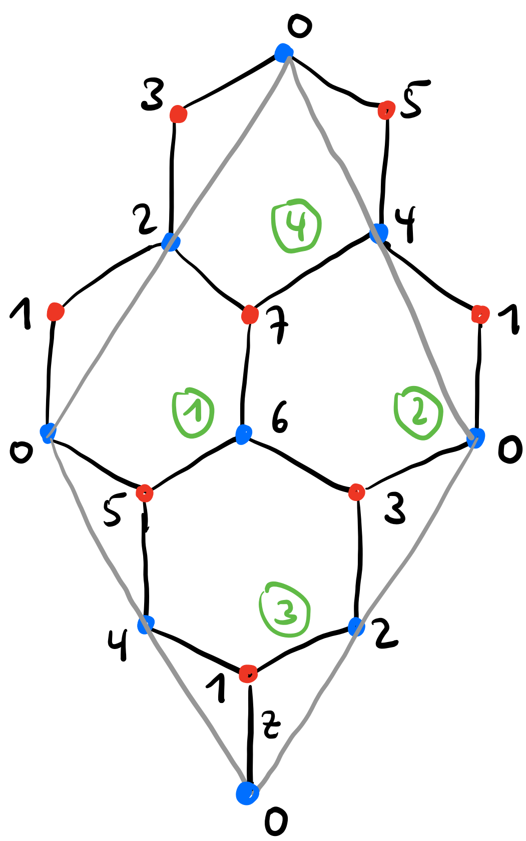
$$\Rightarrow C_G(\vec{\theta}) = 1 - p_0$$
$$C_L(\vec{\theta}) = 1 - \frac{1}{4}\left[4 p_0 + 3 p_2 + 2 p_5 + 3 p_8 + p_{14}\right]$$

$\Rightarrow$ C_L is more robust to small errors / fluctuations than C_G (more forgiving).

Example:

Kitaev honeycomb model

$$H_K = J \sum_{\langle i,j \rangle_\alpha} S_i^\alpha S_j^\alpha$$



Sigmax()

qeye()

tensor([sigmax(), qeye(),...])

$*$ $\hat{=}$ tensor product of Qobjects

Qobj.full()

= numpy.array

Common ansatz strategies:

1. • hardware efficient ansatz (use native gates of QPU)

2. • model specific ansatz, e.g. Hamiltonian variational ansatz (HVA): $H = \sum_i a_i h_i = H_1 + \ldots + H_m$

$$U(\vec{\theta}) = \prod_{\underset{\uparrow \text{ layers}}{l=1}}^{\ell} \left[U_{l,m}(\vec{\theta}_{l,m}) \cdots U_{l,1}(\vec{\theta}_{l,1}) \right]$$

For Kitaev model in magnetic field:

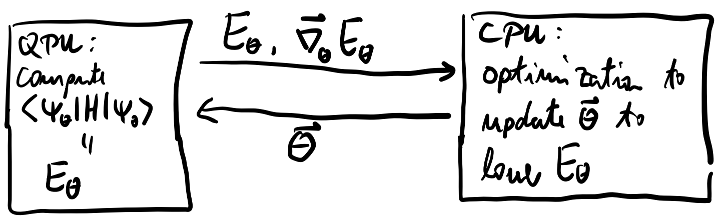
$$U(\vec{\theta}) = \prod_{l=1}^{\ell} \Big[e^{-i\sum_i \theta_{l,6,i} Z_i} e^{-i\sum_i \theta_{l,5,i} Y_i} e^{-i\sum_i \theta_{l,4,i} X_i} \cdot e^{-i\sum_{\langle i,j\rangle_z} \theta_{l,3,i} Z_i Z_j} e^{-i\sum_{\langle i,j\rangle_y} \theta_{l,2,i} Y_i Y_j} e^{-i\sum_{\langle i,j\rangle_x} \theta_{l,1,i} X_i X_j} \Big]$$

Often we also set all $\theta_{l,\alpha,i} = \theta_{l,\alpha}$ $\forall i$ to reduce number of parameters.

3. • adaptive ansatz strategy: grow ansatz dynamically during optimization: ADAPT-VQE: append ansatz by an operator from a predefined pool based on maximizing energy gradient.

Classical optimization:

Classical-quantum feedback loop



Optimization strategies:

- gradient-based:
  - (stochastic) gradient descent
  - quantum natural gradient descent ($\hat{=}$ imaginary time evolution)
  - ADAM
  - BFGS (nonlinear optimization with constraints)
  - SPSS
- non-gradient based:
  - Nelder-Mead
  - BOBYQA
  - Evolutionary algorithms like CMA-ES

Parameter shift rule to compute gradients.

For ansatze $U(\vec{\theta}) = \prod_{i=1}^{N} e^{-i\theta_i P_i}$ with Pauli string P_i,
(like HVA)
(can be relaxed to any hermitian operator with just two eigenvalues)

one can use the parameter shift rule to analytically compute gradients (Mitarai et al. (2018), Schuld et al. (2018)):

$$e^{-i\theta_i P_i} = \cos(\theta_i) - i P_i \sin(\theta_i)$$

$$\Rightarrow \frac{\partial}{\partial \theta_i} U(\vec{\theta}) = U_{j>i}(\vec{\theta}_{j>i}) \left[-\sin(\theta_i) - i P_i \cos\theta_i \right] \cdot U_{j<i}(\vec{\theta}_{j<i})$$

The energy is given by

$$E_\theta = \langle \psi_\theta | H | \psi_\theta \rangle =$$

$$= \langle \psi_0 | U_1^\dagger \cdots U_N^\dagger \, H \, U_N \cdots U_1 | \psi_0 \rangle =$$

$$= \langle \psi_{i-1} | U_i^\dagger \underbrace{\tilde{H}_{i+1}}_{= U_{i+1}^\dagger \cdots U_N^\dagger H U_N \cdots U_{i+1}} U_i \underbrace{| \psi_{i-1} \rangle}_{= U_{i-1} \cdots U_1 | \psi_0 \rangle}$$

Now:

$$\frac{\partial}{\partial \theta_i} E_\theta = i \langle \psi_{i-1} | U_i^\dagger \left[P_i, \tilde{H}_{i+1} \right] U_i | \psi_{i-1} \rangle$$

$$\frac{\partial}{\partial \theta_i} \underbrace{e^{-i\theta_i P_i}}_{= U_i} = -i P_i \, e^{-i\theta_i P_i} = -i e^{-i\theta_i P_i} P_i$$

<u>Now:</u>

$$[P_i, \tilde{H}_{i+1}] = P_i \tilde{H}_{i+1} - \tilde{H}_{i+1} P_i =$$

$$= -i\left(\frac{1}{2}[1 + iP_i]\,\tilde{H}_{i+1}[1 - iP_i] - \frac{1}{2}[1 - iP_i]\,\tilde{H}_{i+1}[1 + iP_i]\right) =$$

$$e^{\mp i\frac{\pi}{4}P_i} = \frac{1}{\sqrt{2}}(1 \mp iP_i)$$

$$= -i\,\frac{1}{2}\left[2iP_i\tilde{H}_{i+1} - 2i\tilde{H}_{i+1}P_i\right] =$$

$$= [P_i, \tilde{H}_{i+1}].$$

$$\Rightarrow [P_i, \tilde{H}_{i+1}] = -i\left[U_i^\dagger\left(\frac{\pi}{4}\right)\tilde{H}_{i+1}U_i\left(\frac{\pi}{4}\right) - U_i^\dagger\left(-\frac{\pi}{4}\right)\tilde{H}_{i+1}U_i\left(-\frac{\pi}{4}\right)\right]$$

The gradient then becomes:

$$\frac{\partial}{\partial \theta_i}E_\theta = i\langle \psi_{i-1}|\,U_i^\dagger\overbrace{[P_i, \tilde{H}_{i+1}]}^{=-i[\dots]}U_i\,|\psi_{i-1}\rangle =$$

$$= \langle \psi_{i-1}|\,U_i^\dagger\left(\theta_i + \frac{\pi}{4}\right)\tilde{H}_{i+1}U_i\left(\theta_i + \frac{\pi}{4}\right)$$

$$- U_i^\dagger\left(\theta_i - \frac{\pi}{4}\right)\tilde{H}_{i+1}U_i\left(\theta_i - \frac{\pi}{4}\right)|\psi_{i-1}\rangle.$$

Example: Kitaev model

- demonstrate VQE for different optimizers
  (if time permits to prepare, otherwise let students work on it).

Quantum natural gradient descent (QNG):

introduced by Stokes et al. (2019).
Use quantum geometry of variational pure states to accelerate convergence to local minima (compared to vanilla GD). Similar method very popular in ML (natural GD), where the Fisher information matrix plays the role of the quantum metric g_{ij}.

Vanilla gradient descent (to find local minima, use multiple starting points to reach global minimum):

Cost function $C(\theta) = \frac{1}{2} \langle \psi_\theta | H | \psi_\theta \rangle$

$$|\psi_\theta\rangle = \prod_{i=1}^{N} \mathcal{U}_i(\theta_i) \, |\psi_0\rangle$$

$$\vec{\theta} = (\theta_1, \dots, \theta_N)$$

$e^{-i\theta_i P_i}$, can also include gates that are θ independent.

Vanilla gradient descent update rule

$$\vec{\theta}_{t+1} := \vec{\theta}_t - \eta \vec{\nabla} C(\theta_t)$$

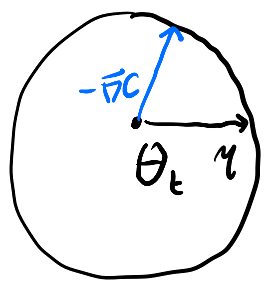
step size

$$= \underset{\vec{\theta}}{\operatorname{argmin}} \left[\langle \vec{\theta} - \vec{\theta}_t , \vec{\nabla} C(\vec{\theta}_t) \rangle + \frac{1}{2\eta} \| \vec{\theta} - \vec{\theta}_t \|_2^2 \right]$$

move in direction of steepest descent

fix step size to be η. Important: $\| \vec{\theta} - \vec{\theta}_t \|_2^2$ is w.r.t. Euclidean l_2-norm in parameter $\vec{\theta}$ space.

$$\| \vec{\theta} - \vec{\theta}_t \|_2^2 = \sum_{i=1}^{N} \left(\theta_i - \theta_{t,i} \right)^2$$



Equivalence of updating rule & argmin:

$$\frac{\partial}{\partial \theta_i} \left[\langle \vec{\theta} - \vec{\theta}_t , \vec{\nabla} C \rangle + \frac{1}{2\eta} \| \vec{\theta} - \vec{\theta}_t \|_2^2 \right] =$$

$$= (\vec{\nabla} C)_i + \frac{1}{\cancel{2}\eta} \cancel{2} (\theta_i - \theta_{t,i}) = 0$$

$$\Rightarrow \quad \theta_i = \theta_{t,i} - \eta (\vec{\nabla} C)_i \qquad \square$$

Now the Euclidean norm in parameter space does not properly reflect the geometry of the states $|\psi_\theta\rangle$. This can slow down convergence.

Fubini-Study metric is unique, unitarily invariant metric on state of pure states $|\psi\rangle\langle\psi| \in \mathbb{C}P^{d-1}$, $d = 2^n$ for qubits:

$$\boxed{d(P_\psi, P_\phi) = \mathrm{acos}(|\langle\psi|\phi\rangle|)} \in [0, \tfrac{\pi}{2}].$$

normalized vectors

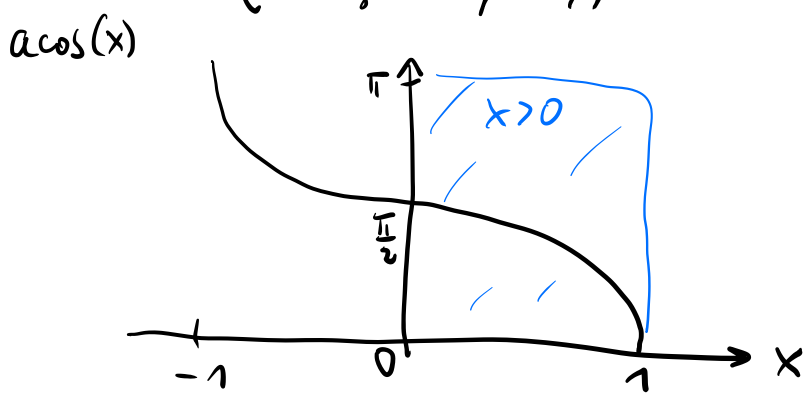
Obeys properties of a metric:

① $d(P_\psi, P_\phi) = 0 \iff P_\psi = P_\phi$

② $d(P_\psi, P_\phi) = d(P_\phi, P_\psi)$

③ $d(P_\psi, P_\chi) \leq d(P_\psi, P_\phi) + d(P_\phi, P_\chi)$.

(triangle inequality)



For our variational states ($\vec{\theta} = (\theta_1, \dots, \theta_N)$):

$$\boxed{d^2(P_\theta, P_{\theta+\partial\theta}) = \sum_{i,j=1}^{N} g_{ij}(\theta)\, \delta\theta_i\, \delta\theta_j}$$

Fubini-Study quantum metric on variational states

Derivation: (Einstein summation)

$$|\psi_{\theta+\partial\theta}\rangle = |\psi_\theta\rangle + \frac{\partial|\psi_\theta\rangle}{\partial\theta_i}\delta\theta_i + \frac{\partial^2|\psi_\theta\rangle}{\partial\theta_j\,\partial\theta_i}\delta\theta_j\,\delta\theta_i + \dots$$

Two useful identities:

$$\langle\psi_\theta|\psi_\theta\rangle = 1$$

$$\Rightarrow \underbrace{\frac{\partial\langle\psi_\theta|}{\partial\theta_i}|\psi_\theta\rangle}_{\text{can also be written as } \left\langle \frac{\partial\psi_\theta}{\partial\theta_i}, \psi_\theta \right\rangle} + \overbrace{\langle\psi_\theta|\frac{\partial\psi_\theta}{\partial\theta_i}\rangle}^{= -i\langle\psi_\theta| i\frac{\partial}{\partial\theta_i}|\psi_\theta\rangle = -i\overset{\mathbb{R}}{A_i}} = 0 \Rightarrow \langle\psi_\theta|\frac{\partial\psi_\theta}{\partial\theta_i}\rangle = i\mathbb{R}.$$

$$\frac{\partial\langle\psi_\theta|}{\partial\theta_i}|\psi_\theta\rangle = \left(\langle\psi_\theta|\frac{\partial\psi_\theta}{\partial\theta_i}\rangle\right)^* = (-iA_i)^* = iA_i.$$

Taking one more derivative:

$$\frac{\partial^2\langle\psi_\theta|}{\partial\theta_j\,\partial\theta_i}|\psi_\theta\rangle + \langle\psi_\theta|\frac{\partial^2\psi_\theta}{\partial\theta_j\,\partial\theta_i}\rangle + \frac{\partial\langle\psi_\theta|}{\partial\theta_i}\frac{\partial|\psi_\theta\rangle}{\partial\theta_j}$$

$$+ \frac{\partial\langle\psi_\theta|}{\partial\theta_j}\frac{\partial|\psi_\theta\rangle}{\partial\theta_i} = 0$$

We want to calculate the distance (squared):

$$\Rightarrow d^2(P_\theta, P_{\theta+\delta\theta}) = \mathrm{acos}^2\left(|\langle\psi_\theta|\psi_{\theta+\delta\theta}\rangle|\right) =$$

We thus need the overlap

$$\langle\psi_\theta|\psi_{\theta+\delta\theta}\rangle = 1 + \langle\psi_\theta|\frac{\partial\psi_\theta}{\partial\theta_i}\rangle\delta\theta_i$$
$$+ \frac{1}{2}\langle\psi_\theta|\frac{\partial^2\psi_\theta}{\partial\theta_j\partial\theta_i}\rangle\delta\theta_i\delta\theta_j$$

Thus,

$$|\langle\psi_\theta|\psi_{\theta+\delta\theta}\rangle|^2 = \langle\psi_{\theta+\delta\theta}|\psi_\theta\rangle\langle\psi_\theta|\psi_{\theta+\delta\theta}\rangle =$$

$$= \left[1 + \langle\frac{\partial\psi_\theta}{\partial\theta_i}|\psi_\theta\rangle\delta\theta_i + \frac{1}{2}\langle\frac{\partial^2\psi_\theta}{\partial\theta_j\partial\theta_i}|\psi_\theta\rangle\delta\theta_i\delta\theta_j\right]$$
$$\cdot\left[1 + \langle\psi_\theta|\frac{\partial\psi_\theta}{\partial\theta_j}\rangle\delta\theta_j + \frac{1}{2}\langle\psi_\theta|\frac{\partial^2\psi_\theta}{\partial\theta_k\partial\theta_\ell}\rangle\delta\theta_k\delta\theta_\ell\right]$$

$$= 1 + \underbrace{\left[\langle\frac{\partial\psi_\theta}{\partial\theta_i}|\psi_\theta\rangle + \langle\psi_\theta|\frac{\partial\psi_\theta}{\partial\theta_i}\rangle\right]}_{=0}\delta\theta_i$$

$$+\Big(\underbrace{\langle \frac{\partial\psi_\theta}{\partial\theta_i}|\psi_\theta\rangle}_{=iA_i}\underbrace{\langle\psi_\theta|\frac{\partial\psi_\theta}{\partial\theta_j}\rangle}_{=-iA_j}+\frac{1}{2}\langle\frac{\partial^2\psi_\theta}{\partial\theta_j\partial\theta_i}|\psi_\theta\rangle$$

$$+\frac{1}{2}\langle\psi_\theta|\frac{\partial^2\psi_\theta}{\partial\theta_j\partial\theta_i}\rangle\Big)\,\partial\theta_i\partial\theta_j =$$

$$=1+\Big[\langle\frac{\partial\psi_\theta}{\partial\theta_i}|\psi_\theta\rangle\langle\psi_\theta|\frac{\partial\psi_\theta}{\partial\theta_j}\rangle$$

$$-\frac{1}{2}\Big(\langle\frac{\partial\psi_\theta}{\partial\theta_j}|\frac{\partial\psi_\theta}{\partial\theta_i}\rangle+\langle\frac{\partial\psi_\theta}{\partial\theta_i}|\frac{\partial\psi_\theta}{\partial\theta_j}\rangle\Big)\Big]\partial\theta_i\partial\theta_j$$

$$+\dots$$

$$\Rightarrow |\langle\psi_\theta|\psi_{\theta+\delta\theta}\rangle| = \sqrt{|\langle\psi_\theta|\psi_{\theta+\partial\theta}\rangle|^2} =$$

$$=1+\frac{1}{2}\langle\frac{\partial\psi_\theta}{\partial\theta_i}|\psi_\theta\rangle\langle\psi_\theta|\frac{\partial\psi_\theta}{\partial\theta_j}\rangle\,\partial\theta_i\partial\theta_j$$

$$-\frac{1}{4}\Big(\langle\frac{\partial\psi_\theta}{\partial\theta_j}|\frac{\partial\psi_\theta}{\partial\theta_i}\rangle+\langle\frac{\partial\psi_\theta}{\partial\theta_i}|\frac{\partial\psi_\theta}{\partial\theta_j}\rangle\Big)\,\partial\theta_i\partial\theta_j$$

Using that $\operatorname{acos}^2(1-x) = 2x + O(x^2)$, we finally obtain

$$d^2(P_\theta, P_{\theta+\delta\theta}) = \mathrm{acos}^2\left(|\langle\psi_\theta|\psi_{\theta+\delta\theta}\rangle|\right) =$$

$$= \Bigg[\overbrace{\frac{1}{2}\left(\left\langle \frac{\partial\psi_\theta}{\partial\theta_j}\middle|\frac{\partial\psi_\theta}{\partial\theta_i}\right\rangle + \left\langle \frac{\partial\psi_\theta}{\partial\theta_i}\middle|\frac{\partial\psi_\theta}{\partial\theta_j}\right\rangle\right)}^{= \mathrm{Re}\left(\left\langle \frac{\partial\psi_\theta}{\partial\theta_j}\middle|\frac{\partial\psi_\theta}{\partial\theta_i}\right\rangle\right)}$$

$$- \underbrace{\left\langle \frac{\partial\psi_\theta}{\partial\theta_i}\middle|\psi_\theta\right\rangle\left\langle\psi_\theta\middle|\frac{\partial\psi}{\partial\theta_j}\right\rangle}_{\text{also purely real as } \langle\psi_\theta|\frac{\partial\psi}{\partial\theta_j}\rangle = -iA_j = i\mathbb{R} \text{ is purely imaginary.}} \Bigg] \delta\theta_i\, \delta\theta_j$$

$$\Rightarrow d^2(P_\theta, P_{\theta+\delta\theta}) = \mathrm{Re}\Bigg[\left\langle \frac{\partial\psi_\theta}{\partial\theta_j}\middle|\frac{\partial\psi_\theta}{\partial\theta_i}\right\rangle$$

$$- \left\langle \frac{\partial\psi_\theta}{\partial\theta_i}\middle|\psi_\theta\right\rangle\left\langle\psi_\theta\middle|\frac{\partial\psi_\theta}{\partial\theta_j}\right\rangle\Bigg]\delta\theta_i\,\delta\theta_j =$$

$$= \mathrm{Re}\big[\underbrace{G_{ij}(\theta)}_{\text{Quantum geometric tensor}}\big]\,\delta\theta_i\,\delta\theta_j = \underbrace{g_{ij}(\theta)}_{\text{Fubini-Study metric}}\,\delta\theta_i\,\delta\theta_j .$$

Note that $\mathrm{Im}(G_{ij}) = \frac{1}{2i}\left(\left\langle \frac{\partial \psi_\theta}{\partial \theta_j} \middle| \frac{\partial \psi_\theta}{\partial \theta_i} \right\rangle - \left\langle \frac{\partial \psi_\theta}{\partial \theta_i} \middle| \frac{\partial \psi_\theta}{\partial \theta_j} \right\rangle \right)$

$$= -\frac{i}{2}\left[\frac{\partial}{\partial \theta_j} \underbrace{\left\langle \psi_\theta \middle| \frac{\partial \psi_\theta}{\partial \theta_i} \right\rangle}_{= -i A_i \text{ (Berry connection)}} - \frac{\partial}{\partial \theta_i} \left\langle \psi_\theta \middle| \frac{\partial \psi_\theta}{\partial \theta_j} \right\rangle \right] =$$

$$= -\frac{1}{2}\left[\frac{\partial}{\partial \theta_j} A_i - \frac{\partial}{\partial \theta_i} A_j \right]$$ = curl of Berry connection = Berry curvature.

<u>Recap:</u>

The Fubini-Study metric describes distances of pure states.
It is unitarily invariant, i.e., this distance measure is invariant under reparametrizations of $|\psi_\theta\rangle$.

<u>Quantum natural gradient descent</u> (QNG):

$$\theta_{t+1} := \underset{\theta}{\mathrm{argmin}} \left[\langle \theta - \theta_t , \vec{\nabla} C(\theta) \rangle + \frac{1}{2\eta} \| \theta - \theta_t \|^2_{g(\theta)} \right]$$

step size determined using Fubini-Study (FS) metric

First-order optimality:

$$\frac{\partial}{\partial \theta_i}\left[\langle \theta - \theta_t , \vec{\nabla} C(\theta)\rangle + \frac{1}{2\eta} \overbrace{\| \theta - \theta_t \|^2_{g(\theta)}}^{= \langle \theta - \theta_t , g(\theta - \theta_t)\rangle}\right] = 0$$

$$\Leftrightarrow (\nabla C(\theta))_i + \frac{1}{2\eta}\left[g_{ij}(\theta_j - \theta_{t,j}) + \underbrace{(\theta_j - \theta_{t,j}) g_{ji}}_{\substack{g_{ij} = g_{ji} \\ \Rightarrow \text{ equal to} \\ g_{ij}(\theta_j - \theta_{t,j})}}\right] = 0$$

$$\Rightarrow g_{ij}(\theta_t)\left(\theta_i - \theta_{t,i}\right) = -\eta\left[\vec{\nabla} C(\theta_t)\right]_i$$

$$\boxed{\begin{aligned} &\Leftrightarrow\; g(\theta_t)\left(\theta_{t+1} - \theta_t\right) = -\eta\, \vec{\nabla} C(\theta_t) \\ &\Leftrightarrow\; \theta_{t+1} = \theta_t - \eta\, \bar{g}^{-1}(\theta_t)\, \vec{\nabla} C(\theta_t) \end{aligned}}$$

Updating rules for QNG

generalized inverse

$$[g + \xi \mathbb{1}]^{-1} = \bar{g}^{-1}$$

with $\xi \ll 1$.

Relation to quantum imaginary time evolution:

It turns out that QNG update rule corresponds to quantum imaginary time evolution (QITE), projected onto the variational manifold and in the infinitesimal step size limit.

Derivation:

QITE is defined by $\left[|\psi_{\bar\theta}\rangle = e^{-H\delta\tau}|\psi_\theta\rangle\right]$:

$$\underset{\delta\theta\in\mathbb{R}^{N_\theta}}{\operatorname{argmin}} \underbrace{\left\| |\psi_{\bar\theta}\rangle - |\psi_{\theta+\delta\theta}\rangle\langle\psi_{\theta+\delta\theta}|\psi_{\bar\theta}\rangle \right\|^2}_{\substack{= \langle\psi_{\bar\theta}|\psi_{\bar\theta}\rangle + \langle\psi_{\theta+\delta\theta}|\psi_{\bar\theta}\rangle \\ -2|\langle\psi_{\theta+\delta\theta}|\psi_{\bar\theta}\rangle|^2}} =$$

minimize by maximizing the last term.

$$= \underset{\delta\theta\in\mathbb{R}^{N_\theta}}{\operatorname{argmax}} \left|\langle\psi_{\bar\theta}|\psi_{\theta+\delta\theta}\rangle\right|^2 .$$

Expand expression to quadratic order in $\delta\theta$, $\delta\tau$ and take 1st order optimality condition.

First we find

$$\langle \psi_{\bar\theta} | \psi_{\theta+\delta\theta} \rangle = \langle \psi_{\bar\theta} | \psi_\theta \rangle + \langle \psi_{\bar\theta} | \frac{\partial \psi_\theta}{\partial \delta\theta_i} \rangle \delta\theta_i$$

$$+ \frac{1}{2} \langle \psi_{\bar\theta} | \frac{\partial^2 \psi_\theta}{\partial \delta\theta_j \, \partial \delta\theta_i} \delta\theta_j \, \delta\theta_i + \dots$$

$$\Rightarrow \left| \langle \psi_{\bar\theta} | \psi_{\theta+\delta\theta} \rangle \right|^2 =$$

$$= \left[\langle \psi_{\bar\theta} | \psi_\theta \rangle + \langle \psi_{\bar\theta} | \frac{\partial \psi_\theta}{\partial \delta\theta_i} \rangle \delta\theta_i + \frac{1}{2} \langle \psi_{\bar\theta} | \frac{\partial^2 \psi_\theta}{\partial \delta\theta_j \, \partial \delta\theta_i} \delta\theta_j \, \delta\theta_i \right]$$

$$\cdot \left[\langle \psi_\theta | \psi_{\bar\theta} \rangle + \langle \frac{\partial \psi_\theta}{\partial \delta\theta_i} | \psi_{\bar\theta} \rangle \delta\theta_i + \frac{1}{2} \langle \frac{\partial^2 \psi_\theta}{\partial \delta\theta_j \, \partial \delta\theta_i} | \psi_{\bar\theta} \rangle \delta\theta_j \, \delta\theta_i \right]$$

$$= \underbrace{\left| \langle \psi_{\bar\theta} | \psi_\theta \rangle \right|^2}_{= \langle \psi_{\bar\theta} | \psi_\theta \rangle \langle \psi_\theta | \psi_{\bar\theta} \rangle} + \left[\langle \psi_{\bar\theta} | \psi_\theta \rangle \langle \frac{\partial \psi_\theta}{\partial \delta\theta_i} | \psi_{\bar\theta} \rangle \right.$$

$$\left. + \langle \psi_\theta | \psi_{\bar\theta} \rangle \langle \psi_{\bar\theta} | \frac{\partial \psi_\theta}{\partial \delta\theta_i} \rangle \right] \delta\theta_i$$

$$+ \left[\langle \psi_{\bar\theta} | \frac{\partial \psi_\theta}{\partial \delta\theta_i} \rangle \langle \frac{\partial \psi_\theta}{\partial \delta\theta_j} | \psi_{\bar\theta} \rangle \right.$$

$$\left. + \langle \psi_{\bar\theta} | \psi_\theta \rangle \frac{1}{2} \langle \frac{\partial^2 \psi_\theta}{\partial \delta\theta_j \partial \delta\theta_i} | \psi_{\bar\theta} \rangle + \langle \psi_\theta | \psi_{\bar\theta} \rangle \frac{1}{2} \langle \psi_{\bar\theta} | \frac{\partial^2 \psi_\theta}{\partial \delta\theta_i \, \partial \delta\theta_j} \rangle \right]$$

$$\cdot\, \delta\theta_i\, \delta\theta_j \,.$$

Now expand $|\psi_{\bar\theta}\rangle = e^{-H\delta\tau}|\psi_\theta\rangle = (1 - H\delta\tau)|\psi_\theta\rangle$

and keep terms up to second order in $\delta\tau, \delta\theta_i$:

$$|\langle\psi_{\bar\theta}|\psi_{\theta+\delta\theta}\rangle|^2 = |\langle\psi_{\bar\theta}|\psi_\theta\rangle|^2$$

does not depend on $\delta\theta_i$, so drops out when computing 1st order optimality $\frac{\partial}{\partial\delta\theta_i}(\,..\,) = 0$

$$+ \left(\langle\psi_\theta|\psi_\theta\rangle - 2\delta\tau\langle\psi_\theta|H|\psi_\theta\rangle\right)\Big[\underbrace{\langle\frac{\partial\psi_\theta}{\partial\delta\theta_i}|\psi_\theta\rangle + \langle\psi_\theta|\frac{\partial\psi_\theta}{\partial\delta\theta_i}\rangle}_{=0}\Big]\delta\theta_i$$

$$- \underbrace{\Big[\langle\frac{\partial\psi_\theta}{\partial\delta\theta_i}|H|\psi_\theta\rangle + \langle\psi_\theta|H|\frac{\partial\psi}{\partial\delta\theta_i}\rangle\Big]}_{= \frac{\partial}{\partial\delta\theta_i}\langle\psi_\theta|H|\psi_\theta\rangle = \frac{\partial}{\delta\theta_i}E_\theta}\delta\tau\,\delta\theta_i$$

$$+ \Big[\langle\psi_\theta|\frac{\partial\psi_\theta}{\partial\delta\theta_i}\rangle\langle\frac{\partial\psi_\theta}{\partial\delta\theta_j}|\psi_\theta\rangle - \frac{1}{2}\Big(\langle\frac{\partial\psi_\theta}{\partial\delta\theta_i}|\frac{\partial\psi_\theta}{\partial\delta\theta_j}\rangle$$

$$+ \langle\frac{\partial\psi_\theta}{\partial\delta\theta_j}|\frac{\partial\psi_\theta}{\partial\delta\theta_i}\rangle\Big)\Big]\delta\theta_i\,\delta\theta_j$$

$$= - \mathrm{Re}[G_{ij}]\, \delta\theta_i\, \delta\theta_j$$

$$\Leftrightarrow |\langle \psi_{\bar\theta} | \psi_{\theta+\delta\theta}\rangle|^2 = |\langle \psi_{\bar\theta} | \psi_\theta\rangle|^2 - \frac{\partial}{\partial \delta\theta_i} \langle \psi_\theta | H | \psi_\theta \rangle \delta\theta_i\, \delta\tau - g_{ij}(\theta)\, \delta\theta_i\, \delta\theta_j$$

To find $\operatorname{argmax}_{\delta\theta} |\langle \psi_{\bar\theta} | \psi_{\theta+\delta\theta}\rangle|^2$, we look at first order optimality

$$\frac{\partial}{\partial \delta\theta_i} |\langle \psi_{\bar\theta} | \psi_{\theta+\delta\theta}\rangle|^2 = 0$$

$$\Leftrightarrow - \frac{\partial}{\partial \delta\theta_i} \langle \psi_\theta | H | \psi_\theta \rangle \delta\tau - 2 g_{ij}\, \delta\theta_j = 0$$

$$\Leftrightarrow g_{ij}(\theta)\, \delta\theta_j = - \frac{\partial}{\partial \delta\theta_i} \underbrace{\tfrac{1}{2} \langle \psi_\theta | H | \psi_\theta \rangle}_{= C(\theta)} \delta\tau$$

QITE

$$\Rightarrow \boxed{g_{ij}(\theta)\, \delta\theta_j = - \frac{\partial}{\partial \delta\theta_i} C(\theta)\, \delta\tau}$$

This is exactly the QNG update rule

$\Rightarrow$ $\boxed{\text{QITE} \hat{=} \text{QNG}}$

Using vector notation, we find

$$\Rightarrow \quad g\,\delta\theta = -\vec{\nabla} C(\theta)\,\delta\tau$$

In the limit $\delta\tau \to 0$, we find

$$\boxed{g[\theta(\tau)]\,\dot{\theta}(\tau) = -\vec{\nabla} C[\theta(\tau)]} .$$

This corresponds to the VQITE EOM derived from McLachlan's principle as we show next.

McLachlan's variational principle (Li, Benjamin et al, (2018))

Density matrix under imaginary time evolution:

$$\rho(\tau) = \frac{e^{-H\tau}\rho(0)\,e^{-H\tau}}{\mathrm{Tr}\left[e^{-2H\tau}\rho(0)\right]}$$

$$\Rightarrow \frac{\partial\rho}{\partial\tau} = -H\rho(\tau) - \rho(\tau)H - \overbrace{\frac{e^{-H\tau}\rho(0)\,e^{-H\tau}}{\mathrm{Tr}\left[e^{-2H\tau}\rho(0)\right]}}^{=\rho(\tau)}\;\overbrace{\frac{\mathrm{Tr}\left[-2He^{-2H\tau}\rho(0)\right]}{\mathrm{Tr}\left[e^{-2H\tau}\rho(0)\right]}}^{=-2\,\mathrm{Tr}[H\rho(\tau)]}$$

$$\Longleftrightarrow \boxed{\frac{\partial \rho}{\partial \tau} = -\{H, \rho(\tau)\} + 2\langle H\rangle_{\rho(\tau)}\, \rho(\tau)}$$

Imag. time evolution of density matrix

Parametrize DM $\rho[\theta(\tau)]$ and then derive equation of motion of variational parameters using the variational principle.

What we want to minimize is the (McLachlan) distance between the time evolution of the variational parameters and the exact imag. time evolution:

$\|\rho\|^2 = \mathrm{Tr}[\rho^\dagger \rho]$ (Frobenius norm of matrices)

$$L^2 = \left\| \frac{\partial \rho[\theta(\tau)]}{\partial \theta_i} \dot{\theta}_i - \left(\frac{\partial \rho}{\partial \tau}\right)_{exact} \right\|^2 =$$

$$= \left\| \frac{\partial \rho[\theta(\tau)]}{\partial \theta_i} \dot{\theta}_i + \{H, \rho(\tau)\} - 2\langle H\rangle \rho(\tau) \right\|^2 =$$

$$= \mathrm{Tr}\left\{\left[\left(\frac{\partial \rho}{\partial \theta_i}\right)^\dagger \dot{\theta}_i + \{H, \rho\} - 2\langle H\rangle \rho\right]\left[\left(\frac{\partial \rho}{\partial \theta_j}\right)\dot{\theta}_j + \{H, \rho\} - 2\langle H\rangle \rho\right]\right\} =$$

$\rho^\dagger = \rho$, $H^\dagger = H$

$\theta \in \mathbb{R}$

$$= \mathrm{Tr}\left[\left(\frac{\partial \rho}{\partial \theta_i}\right)^{\dagger}\left(\frac{\partial \rho}{\partial \theta_j}\right)\dot{\theta}_i\,\dot{\theta}_j\right]$$

$$+ \mathrm{Tr}\Big[\left(\frac{\partial \rho}{\partial \theta_i}\right)^{\dagger}\left(H\rho + \rho H - 2\langle H\rangle \rho\right)$$

$$+ \left(H\rho + \rho H - 2\langle H\rangle\rho\right)\left(\frac{\partial\rho}{\partial\theta_i}\right)\Big]\,\dot{\theta}_i$$

$$+ \mathrm{Tr}\left[\left(H\rho + \rho H - 2\langle H\rangle\rho\right)\left(H\rho + \rho H - 2\langle H\rangle\rho\right)\right].$$

We consider pure states in the following:

$$\rho(\tau) = |\psi(\tau)\rangle\langle\psi(\tau)| .$$

Then,

$$① \ \mathrm{Tr}\left[\left(\frac{\partial \rho}{\partial \theta_i}\right)^{\dagger}\overbrace{\left(\frac{\partial \rho}{\partial \theta_j}\right)}^{= \frac{\partial}{\partial\theta_j}|\psi(\tau)\rangle\langle\psi(\tau)| = \left|\frac{\partial\psi(\tau)}{\partial\theta_j}\right\rangle\langle\psi(\tau)| + |\psi(\tau)\rangle\left\langle\frac{\partial\psi(\tau)}{\partial\theta_j}\right|}\dot{\theta}_i\,\dot{\theta}_j\right] =$$

$$= \mathrm{Tr}\left[\left(|\psi\rangle\left\langle\frac{\partial\psi}{\partial\theta_i}\right| + \left|\frac{\partial\psi}{\partial\theta_i}\right\rangle\langle\psi|\right)\left(\left|\frac{\partial\psi}{\partial\theta_j}\right\rangle\langle\psi| + |\psi\rangle\left\langle\frac{\partial\psi}{\partial\theta_j}\right|\right)\right.$$

$$\dot{\theta}_i\,\dot{\theta}_j =$$

$$= \Big[\langle \frac{\partial \psi}{\partial \theta_i} | \frac{\partial \psi}{\partial \theta_j} \rangle + \overbrace{\langle \frac{\partial \psi}{\partial \theta_i} | \psi \rangle}^{= iA_i} \overbrace{\langle \frac{\partial \psi}{\partial \theta_j} | \psi \rangle}^{= iA_j} \rightarrow -A_i A_j$$

$$+ \underbrace{\langle \psi | \frac{\partial \psi}{\partial \theta_j} \rangle}_{= -i A_j} \underbrace{\langle \psi | \frac{\partial \psi}{\partial \theta_i} \rangle}_{-i A_i} \rightarrow -A_i A_j + \langle \frac{\partial \psi}{\partial \theta_j} | \frac{\partial \psi}{\partial \theta_i} \rangle \Big] \dot{\theta}_i \dot{\theta}_j$$

$$= 2 \,\mathrm{Re} \Big[\underbrace{\langle \frac{\partial \psi}{\partial \theta_i} | \frac{\partial \psi}{\partial \theta_j} \rangle + \langle \psi | \frac{\partial \psi}{\partial \theta_i} \rangle \langle \psi | \frac{\partial \psi}{\partial \theta_j} \rangle}_{G_{ij}[\theta(\tau)]} \Big] \cdot \dot{\theta}_i \dot{\theta}_j$$

$$= 2 \,\mathrm{Re} \big[G_{ij}[\theta(\tau)] \big] \dot{\theta}_i \dot{\theta}_j = 2\, g_{ij}[\theta(\tau)] \dot{\theta}_i \dot{\theta}_j .$$

(2)

$$\mathrm{Tr} \Big[\overbrace{\Big(\frac{\partial \rho}{\partial \theta_i} \Big)^{+}}^{= |\frac{\partial \psi}{\partial \theta_i}\rangle\langle\psi| + |\psi\rangle\langle\frac{\partial \psi}{\partial \theta_i}|} \big(\underbrace{H \overset{=|\psi\rangle\langle\psi|}{\rho} + \rho H}_{\text{in principle } \rho H + H\rho} - 2 \langle H \rangle \rho \big)$$

focus on pure states

$$+ \big(H\rho + \rho H - 2\langle H \rangle \rho \big) \underbrace{\Big(\frac{\partial \rho}{\partial \theta_i} \Big)}_{= |\frac{\partial \psi}{\partial \theta_i}\rangle\langle\psi| + |\psi\rangle\langle\frac{\partial \psi}{\partial \theta_i}|} \Big] \dot{\theta}_i =$$

see below (1)+(2)+(3)+(4)

$$\hat{=}\; 2 \big(\underbrace{\langle \frac{\partial \psi}{\partial \theta_i} | H | \psi \rangle + \langle \psi | H | \frac{\partial \psi}{\partial \theta_i} \rangle}_{= 2\,\mathrm{Re} (\langle \frac{\partial \psi}{\partial \theta_i} | H | \psi \rangle)}$$

$$-4\langle H\rangle \Big(\underbrace{\langle\psi|\frac{\partial\psi}{\partial\theta_i}\rangle + \langle\frac{\partial\psi}{\partial\theta_i}|\psi\rangle}_{=0}\Big)\Big\}\,\dot{\theta}_i =$$

$$= 4\,\mathrm{Re}\left(\langle\frac{\partial\psi}{\partial\theta_i}|H|\psi\rangle\right).$$

3

$$\mathrm{Tr}\left[\left(H\rho+\rho H-2\langle H\rangle\rho\right)\left(H\rho+\rho H-2\langle H\rangle\rho\right)\right] =$$

$$= \cancel{\langle H\rangle^2} + \underline{\langle H^2\rangle} - 2\cancel{\langle H\rangle^2} + \underline{\langle H^2\rangle} + \cancel{\langle H\rangle^2} - 2\langle H\rangle^2$$

$$-2\cancel{\langle H\rangle^2} - 2\cancel{\langle H\rangle^2} + 4\cancel{\langle H\rangle^2} =$$

$$= 2\left(\langle H^2\rangle - \langle H\rangle^2\right).$$

Now, we can collect the three terms

$$\boxed{\begin{aligned} L^2 &= 1 + 2 + 3 = 2\,\mathrm{Re}\left[G_{ij}[\theta(\tau)]\right]\dot{\theta}_i\dot{\theta}_j \\ &+ 4\,\mathrm{Re}\left(\langle\frac{\partial\psi}{\partial\theta_i}|H|\psi\rangle\right)\dot{\theta}_i + 2\left(\langle H^2\rangle - \langle H\rangle^2\right)\end{aligned}}$$

Extra calculation for step (2) above:

(1)
$$\mathrm{Tr}\left[\left(\frac{\partial \rho}{\partial \theta_i}\right)^{+} H\rho\right] = \mathrm{Tr}\left[\left(\left|\frac{\partial \psi}{\partial \theta_i}\right\rangle\langle\psi| + |\psi\rangle\left\langle\frac{\partial \psi}{\partial \theta_i}\right|\right) H |\psi\rangle\langle\psi|\right]$$

$$= \left\langle\psi\middle|\frac{\partial \psi}{\partial \theta_i}\right\rangle \langle H\rangle + \left\langle\frac{\partial \psi}{\partial \theta_i}\middle| H \middle|\psi\right\rangle .$$

(2)
$$\mathrm{Tr}\left[\left(\frac{\partial \rho}{\partial \theta_i}\right)^{+} \rho H\right] = \mathrm{Tr}\left[\left(\left|\frac{\partial \psi}{\partial \theta_i}\right\rangle\langle\psi| + |\psi\rangle\left\langle\frac{\partial \psi}{\partial \theta_i}\right|\right) |\psi\rangle\langle\psi| H\right] =$$

$$= \left\langle\psi\middle| H \middle|\frac{\partial \psi}{\partial \theta_i}\right\rangle + \underbrace{\left\langle\frac{\partial \psi}{\partial \theta_i}\middle|\psi\right\rangle}_{= -\left\langle\psi\middle|\frac{\partial \psi}{\partial \theta_i}\right\rangle} \langle H\rangle$$

$$\underbrace{\left|\frac{\partial \psi}{\partial \theta_i}\right\rangle\langle\psi| + |\psi\rangle\left\langle\frac{\partial \psi}{\partial \theta_i}\right|}$$

(3)
$$\mathrm{Tr}\left[H\rho \frac{\partial \rho}{\partial \theta_i}\right] = \langle H\rangle \left\langle\psi\middle|\frac{\partial \psi}{\partial \theta_i}\right\rangle + \left\langle\frac{\partial \psi}{\partial \theta_i}\middle| H \middle|\psi\right\rangle$$

(4)
$$\mathrm{Tr}\left[\rho H \frac{\partial \rho}{\partial \theta_i}\right] = \left\langle\psi\middle| H \middle|\frac{\partial \psi}{\partial \theta_i}\right\rangle + \langle H\rangle \underbrace{\left\langle\frac{\partial \psi}{\partial \theta_i}\middle|\psi\right\rangle}_{= -\left\langle\psi\middle|\frac{\partial \psi}{\partial \theta_i}\right\rangle}$$

It is now straightforward to derive an equation of motion for the variational parameters $\theta(\tau)$ that minimizes L^2 using the variational principle:

$$L^2 = 2\,\mathrm{Re}[G_{ij}]\,\dot{\theta}_i\dot{\theta}_j + 4\,\underbrace{\mathrm{Re}\left(\left\langle \frac{\partial\psi}{\partial\theta_i}\middle|H\middle|\psi\right\rangle\right)}_{=\frac{1}{2}\frac{\partial}{\partial\theta_i}\langle\psi|H|\psi\rangle}\dot{\theta}_i + 2\left(\langle H^2\rangle - \langle H\rangle^2\right)$$

$$\delta L^2 = 0 \Rightarrow -\frac{d}{d\tau}\frac{\partial L}{\partial\dot{\theta}_i} + \frac{\partial L}{\partial\theta_i} = 0 \quad \forall i$$

$$\Leftrightarrow 4\,\mathrm{Re}[G_{ij}]\,\dot{\theta}_j + 4\,\mathrm{Re}\left(\left\langle\frac{\partial\psi}{\partial\theta_i}\middle|H\middle|\psi\right\rangle\right) = 0$$

$$\Leftrightarrow g_{ij}\,\dot{\theta}_j = -\frac{\partial}{\partial\theta_i}\underbrace{\frac{1}{2}\langle\psi|H|\psi\rangle}_{=C(\theta)\text{ from earlier}} = -\frac{\partial}{\partial\theta_i}C(\theta)$$

$$\Rightarrow \boxed{g[\theta(\tau)]\,\dot{\theta}(\tau) = -\vec{\nabla}C[\theta(\tau)]}$$

Same equation as we had derived before starting from $\underset{\delta\theta \in \mathbb{R}^{N_\theta}}{\operatorname{argmin}} \left\| |\psi_{\bar\theta}\rangle - |\psi_{\theta+\delta\theta}\rangle\langle\psi_{\theta+\delta\theta}|\psi_{\bar\theta}\rangle \right\|^2 .$

Now that we have a classical EOM for the variational parameters, what do we need to address:

1. measure $g_{ij}[\theta]$ and $\mathrm{Re}\left(\langle \frac{\partial\psi}{\partial\theta_i}|H|\psi\rangle\right)$

   on quantum computers.

YONG-XIN YAO *et al.* PRX QUANTUM **2**, 030307 (2021)

diagonal elements are $\langle A_\mu^2 \rangle$

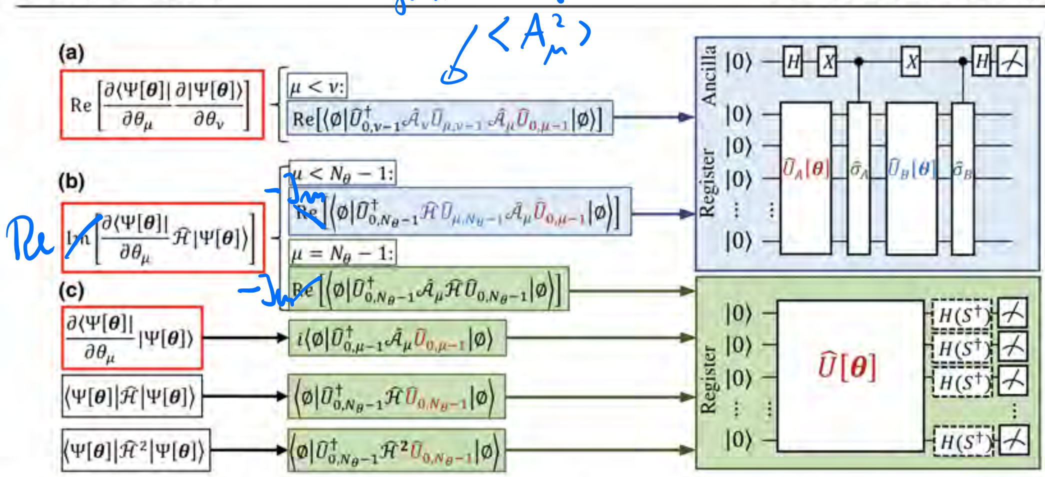


FIG. 2. Quantum-circuit implementation of the AVQDS algorithm. The left column lists the unique terms to be evaluated in Eqs. (5) and (7) of VQDS, with the terms (a)–(c) highlighted in red also involved in the ansatz adaptive procedure in AVQDS. The middle column specifies the expressions when the wave-function ansatz takes the pseudo-Trotter form of Eq. (10) with $\hat{U}_{j,k}[\theta] = \prod_{\mu=j}^{k} e^{-i\theta_\mu \hat{\mathcal{A}}_\mu}$ and $|\Psi_0\rangle = |\emptyset\rangle \equiv \otimes_{j=0}^{N-1} |0\rangle$ for an N-qubit system. Two types of quantum circuits are adopted: a green block for the direct measurement circuit, and a blue block for a generalized Hadamard test circuit [27,45]. The direct measurement circuit includes optional Hadamard gate H or Hadamard-phase gate $HS^\dagger$ when measuring X or Y-Pauli strings present in $\hat{\mathcal{A}}_\mu$, $\hat{\mathcal{A}}_\mu^2$, $\hat{\mathcal{H}}$, and $\hat{\mathcal{H}}^2$. Accord-

Can be done using Hadamard test circuits & direct measurements.

② How to optimally distribute measurements (shots) across different circuits?

- Kozcov, alternatives

③ How to select the ansatz?

- fixed ansatz
- adaptively expanded ansatz

④ How to best deal with inversion of g_{ij}, which is often singular in practice (large condition number)?

- directly solve linear system of equations

$$g \dot{\theta} + \vec{\nabla} C(\theta) = 0$$

e.g. using

numpy.linalg.lstsq

`linalg.lstsq(a, b, rcond='warn')` [source]

Return the least-squares solution to a linear matrix equation.

Computes the vector *x* that approximately solves the equation `a @ x = b`. The equation may be under-, well-, or over-determined (i.e., the number of linearly independent rows of *a* can be less than, equal to, or greater than its number of linearly independent columns). If *a* is square and of full rank, then *x* (but for round-off error) is the "exact" solution of the equation. Else, *x* minimizes the Euclidean 2-norm $||b - ax||$. If there are multiple minimizing solutions, the one with the smallest 2-norm $||x||$ is returned.

Parameters: **a** : *(M, N) array_like*

"Coefficient" matrix.

b : *{(M,), (M, K)} array_like*

Ordinate or "dependent variable" values. If *b* is two-dimensional, the least-squares solution is calculated for each of the *K* columns of *b*.

rcond : *float, optional*

Cut-off ratio for small singular values of *a*. For the purposes of rank determination, singular values are treated as zero if they are smaller than *rcond* times the largest singular value of *a*.

- use Tikhonov regularization

$$g \longrightarrow g + \xi \mathbb{1} \Rightarrow \tilde{g}^{-1} = [g + \xi \mathbb{1}]^{-1}$$

$\xi \ll 1$ (e.g. typically $10^{-1} \le \xi \le 10^{-6}$).

③ How to select ansatz:

Adaptive ansatz generation: flexible and shown to produce shallow circuits with near optimal scaling of N_θ and N_{CNOT} with system size, when optimal scaling is known.

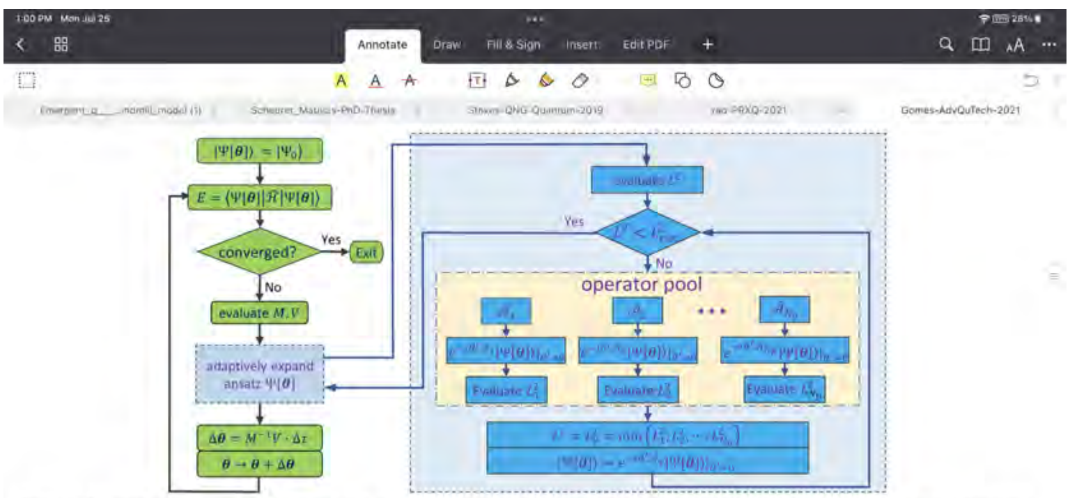


Figure 1. Schematic illustration of variational quantum imaginary time evolution algorithm, with an additional module to adaptively expand the ansatz. The green flowchart on the left shows a typical VQITE calculation. In AVQITE, a module (blue) is introduced to adaptively expand the variational ansatz by selectively appending parametric rotation gates to keep the McLachlan distance L^2 under a threshold L^2_{cut} along the imaginary-time evolution path.

VQITE is formulated in the exactly same form as real time variational quantum dynamics simulations (VQDS),[18,34] with the exception of the definition of vector V in Equation (4) due to the different superoperator $\mathcal{L}[\hat{\rho}]$ in the Liouville–von Neumann-type equation.

2.1.2. Flowchart

A typical VQITE calculation, which integrates the equation of

a product of N_θ multi-qubit rotation gates in a pseudo-Trotter form:

$$|\Psi[\theta]\rangle = \prod_{\mu=1}^{N_\theta} e^{-i\theta_\mu \hat{A}_\mu} |\Psi_0\rangle \tag{8}$$

where $\hat{A}_\mu$ are Hermitian operators. The unitary coupled cluster ansatz and its variants,[6,23,24] the Hamiltonian variational

② How to optimally distribute measurements.

① Koczor et al. suggest to minimize

$$(\Delta\dot\theta)^2 = \sum_{i=1}^{N_\theta} \Delta\dot\theta_i^2 ,$$ where $(\Delta\dot\theta_i)^2$ is the variance

of $\Delta\dot\theta_i = g^{-1}_{ij} \underset{\overset{\uparrow}{\vec\nabla_i C(\theta)}}{V_j}$ over shot distribution.

Koczor work:

$$(\Delta\dot\theta_i)^2 = g^{-1}_{ij} V_j \, g^{-1}_{il} V_l = g^{-1}_{ij} g^{-1}_{il} V_j V_l .$$

Alternative: minimize: $\Delta\dot\theta_i \, g_{ij} \, \Delta\dot\theta_j = g^{-1}_{ik} V_k \, g_{ij} \, g^{-1}_{jl} V_l =$

$$= g^{-1}_{ik} \, \delta_{il} V_k V_l =$$

$$= g^{-1}_{ik} V_k V_i$$

Error propagation of variances:

Use $f(x,y,z)$, then variance of f is related to variances of x,y,z as

$$S_f^2 = \left(\frac{\partial f}{\partial x}\right)^2 S_x^2 + \left(\frac{\partial f}{\partial y}\right)^2 S_y^2 + \left(\frac{\partial f}{\partial z}\right)^2 S_z^2$$

Here, we have assumed that S_x, S_y, S_z are small and independent.

Here:

variance of $\dot{\theta}_i$ due to shots. Note $\dot{\theta}_i = g_{ij}^{-1} V_j$.

$$\Rightarrow \quad S_{\dot{\theta}_i}^2 = \sum_j \left[\frac{\partial}{\partial g_{ij}^{-1}} \left(g_{ik}^{-1} V_k\right)\right]^2 S_{g_{ij}^{-1}}^2$$

$$+ \sum_j \left[\frac{\partial}{\partial V_j} \left(g_{ik}^{-1} V_k\right)\right] S_{V_j}^2$$

$$\frac{\partial}{\partial x}(UU^{-1}) = \frac{\partial}{\partial x}(\mathbb{1}) = 0$$

$$\Rightarrow \frac{\partial U}{\partial x} U^{-1} = -U \frac{\partial U^{-1}}{\partial x}$$

$$\Leftrightarrow \frac{\partial U^{-1}}{\partial x} = -U^{-1} \frac{\partial U}{\partial x} U^{-1}.$$

$$= \sum_j \left[V_j^2 \, S_{g_{ij}^{-1}}^2 + \left(g_{ij}^{-1}\right)^2 S_{V_j}^2\right].$$

Now, we can relate $S_{g_{ij}^{-1}}^2$ to $S_{g_{ij}}^2$ in the following way:

$$s^2_{g^{-1}_{ij}} = \sum_{r,l} \left(\frac{\partial g^{-1}_{ij}}{\partial g_{rl}} \right)^2 s^2_{g_{rl}} =$$

$$= \sum_{r,l} \left(- g^{-1}_{in} \overbrace{\frac{\partial g_{nm}}{\partial g_{rl}}}^{\delta_{rn}\,\delta_{lm}} g^{-1}_{mj} \right)^2 s^2_{g_{rl}} =$$

$$= \sum_{r,l} \left(g^{-1}_{ir} \; g^{-1}_{lj} \right)^2 s^2_{g_{rl}}$$

✓ (according to Ref. [60] of Rozcor).

Thus,

$$s^2_{\dot{\theta}_i} = \sum_j \left[V_j^2 \, s^2_{g^{-1}_{ij}} + \left(g^{-1}_{ij} \right)^2 s^2_{V_j} \right] =$$

$$= \sum_{j,r,l} \left[\left(g^{-1}_{ir} \right)^2 \left(g^{-1}_{lj} \right)^2 V_j^2 \, s^2_{g_{rl}} \right] + \sum_j \left[\left(g^{-1}_{ij} \right)^2 s^2_{V_j} \right]$$

The sum of variances over all elements of $\dot{\theta}_i$ is thus given by

$$S^2_{\dot{\theta}} \equiv \sum_i S^2_{\dot{\theta}_i} =$$

$$= \sum_{i,j,r,l} \left[(g^{-1}_{ir})^2 (g^{-1}_{lj})^2 V_j^2 \, S^2_{g_{rl}} \right] + \sum_{i,j} \left[(g^{-1}_{ij})^2 S^2_{V_j} \right] =$$

$$= \sum_{r,l} a_{rl} \, S^2_{g_{rl}} + \sum_r b_r S_{V_r}$$

with

$$a_{rl} = \sum_{i,j} (g^{-1}_{ir})^2 (g^{-1}_{lj})^2 V_j^2$$

$$b_r = \sum_i (g^{-1}_{ir})^2$$

Now, how should we distribute shots to minimize

$$S_\theta^2 = \sum_{r,l} a_{rl}\, S_{g_{rl}}^2 + \sum_r b_r\, S_{V_r}^2$$

under the constraint of fixed

$$N_{opt} = \sum_{r,l} \underbrace{N_{rl}^{(g)}}_{\text{number of measurements of element } g_{rl}} + \sum_r \underbrace{N_r^{(V)}}_{\text{\# of measurements of element } V_r} .$$

These determine the variances of $S_{g_{rl}}^2$ and $S_{V_r}^2$ as

$$S_{g_{rl}}^2 = \frac{\alpha_{rl}^2}{N_{rl}^{(g)}}, \quad \alpha_{rl}^2 = \tilde{S}_{g_{rl}}^2 \text{ (variance of single measurement)}$$

$$\beta_r^2 = \tilde{S}_{V_r}^2$$

$$S_{V_r}^2 = \frac{\beta_r^2}{N_r^{(V)}}$$

with some N independent parameters $\alpha_{rl}, \beta_r \equiv$ variance of single measurements.

Can use Lagrange multipliers: Define

$$g(N_{rl}^{(g)}, N_r^{(V)}) = \sum_{r,l} N_{rl}^{(g)} + \sum_r N_r^{(V)} - N_{opt}$$

We find the local extrema that fulfill $g = 0$ by solving the set of equations:

$$\frac{\partial S_\theta^2}{\partial N_{rl}^{(g)}} = \lambda \frac{\partial g}{\partial N_{rl}^{(g)}} \qquad (1)$$

$$\frac{\partial^2 S_\theta^2}{\partial N_r^{(v)}} = \lambda \frac{\partial g}{\partial N_r^{(v)}} \qquad (2)$$

$$g = 0 \Leftrightarrow \sum_{r,l} N_{r,l}^{(g)} + \sum_r N_r^{(v)} - N_{opt} = 0 \quad (3)$$

$$(1) \Rightarrow -\frac{\alpha_{rl}^2 a_{rl}}{N_{rl}^{(g)2}} = \lambda \Rightarrow N_{rl}^{(g)} = \sqrt{-\frac{\alpha_{rl}^2 a_{rl}}{\lambda}}$$

$$(2) \Rightarrow -\frac{\beta_r^2 b_r}{N_{rl}^{(v)2}} = \lambda \Rightarrow N_{rl}^{(v)} = \sqrt{-\frac{\beta_r^2 b_r}{\lambda}}$$

$$(3) \sum_{r,l} N_{rl}^{(g)} + \sum_r N_r^{(v)} - N_{opt} = 0$$

$$\overset{(1)(2)}{\Longrightarrow} \sum_{r,l} \sqrt{-\frac{\alpha_{rl}^2 a_{rl}}{\lambda}} + \sum_r \sqrt{\frac{-\beta_r^2 b_r}{\lambda}} - N_{opt} = 0$$

We thus find

$$\sqrt{-\frac{1}{\lambda}}\,\overbrace{\left[\sum_{g,l}\sqrt{\alpha_{gl}^2 a_{gl}}+\sum_{g}\sqrt{\beta_g^2 b_g}\right]}^{=\Sigma} = N_{opt}$$

$$\Rightarrow \boxed{\lambda = -\frac{1}{N_{opt}^2}\underbrace{\left[\sum_{g,l}\sqrt{a_{gl}\alpha_{gl}^2}+\sum_{g}\sqrt{b_g\beta_g^2}\right]^2}_{=\Sigma^2}}$$

Now that we have λ, we can find $N_{gl}^{(g)}$, $N_g^{(v)}$ from (1) & (2):

$$N_{gl}^{(g)} = \sqrt{-\frac{1}{\lambda}}\sqrt{a_{gl}\alpha_{gl}^2} =$$

$$= \frac{N_{opt}}{\Sigma}\sqrt{a_{gl}\alpha_{gl}^2} \Rightarrow \boxed{\frac{N_{gl}^{(g)}}{N_{opt}} = \frac{\sqrt{a_{gl}\alpha_{gl}^2}}{\Sigma}}$$

agrees with Kozcar Eq. (C7)

$$N_g^{(v)} = \sqrt{-\frac{1}{\lambda}}\sqrt{b_g\beta_g^2} \Rightarrow \boxed{\frac{N_g^{(v)}}{N_{opt}} = \frac{\sqrt{b_g\beta_g^2}}{\Sigma}}$$

Finally we can relate N_{opt} to S_θ^2 via

$$S_\theta^2 = \sum_{g,l} \frac{a_{gl}\,\alpha_{gl}^2}{N_{gl}^{(g)}} + \sum_{g} \frac{b_g\,\beta_g^2}{N_g^{(v)}} =$$

$$= \sum_{g,l} \frac{\sqrt{a_{gl}\,\alpha_{gl}^2}\;\Sigma}{N_{opt}} + \sum_{g} \frac{\sqrt{b_g\,\beta_g^2}\;\Sigma}{N_{opt}} =$$

$$= \frac{\Sigma^2}{N_{opt}} \Rightarrow \boxed{N_{opt} = \frac{\Sigma^2}{S_\theta^2}}$$

The optimal shot distribution when aiming for specific S_θ^2 is thus given by

$$\boxed{\begin{aligned} N_{gl}^{(g)} &= \frac{\Sigma}{S_\theta^2}\sqrt{a_{gl}\,\alpha_{gl}^2} \\ N_{g}^{(v)} &= \frac{\Sigma}{S_\theta^2}\sqrt{b_g\,\beta_g^2} \end{aligned}}$$

agrees with Koczor (C1), (C2).

Show Koczor figures as slides.

Real-time evolution:

One can derive similar classical EOM for $\theta(t)$ from McLachlan's principle applied to real-time dynamics. Here, the exact state evolution is given by the von-Neumann eq.:

$$\frac{\partial \rho(t)}{\partial t} = -i\left[H, \rho(t)\right]$$

Applying the variational principle to minimize the McLachlan distance L^2 between variational & exact time evolution yields:

$$\delta L^2 = 0$$

$$\Rightarrow \delta \left\| \frac{\partial \rho[\theta(t)]}{\partial \theta_i} \dot{\theta}_i + i \underbrace{\left[H, \rho[\theta(t)]\right]}_{H\rho - \rho H} \right\|^2 = 0$$

Rewriting L^2 as:

$$L^2 = \mathrm{Tr}\Big\{\Big[\Big(\frac{\partial\rho}{\partial\theta_i}\Big)^{\dagger}\dot\theta_i - i(\rho H - H\rho)\Big]$$

$\updownarrow$ note!

$$\Big[\Big(\frac{\partial\rho}{\partial\theta_j}\Big)\dot\theta_j + i(H\rho - \rho H)\Big]\Big\}$$

$$= \mathrm{Tr}\Big[\Big(\frac{\partial\rho}{\partial\theta_i}\Big)^{\dagger}\Big(\frac{\partial\rho}{\partial\theta_j}\Big)\Big]\dot\theta_i\,\dot\theta_j$$

$$+ i\,\mathrm{Tr}\Big[\Big(\frac{\partial\rho}{\partial\theta_i}\Big)^{\dagger}(H\rho - \rho H)\Big]\dot\theta_i$$

$$- i\,\mathrm{Tr}\Big[(\rho H - H\rho)\Big(\frac{\partial\rho}{\partial\theta_i}\Big)\Big]\dot\theta_i$$

$$+ \mathrm{Tr}\big[(H\rho - \rho H)(H\rho - \rho H)\big]$$

<u>Focusing on pure states:</u> $\rho(t) = |\psi(t)\rangle\langle\psi(t)|$

① $$\mathrm{Tr}\Big[\Big(\underbrace{\frac{\partial\rho}{\partial\theta_i}}_{= |\frac{\partial\psi}{\partial\theta_i}\rangle\langle\psi| + |\psi\rangle\langle\frac{\partial\psi}{\partial\theta_i}|}\Big)^{\dagger}\Big(\frac{\partial\rho}{\partial\theta_j}\Big)\Big] =$$

$$= 2\,\mathrm{Re}[G_{ij}(\theta)] \quad \text{as before}$$

(2)

$$\mathrm{Tr}\Big[\overbrace{\Big(\frac{\partial \rho}{\partial \theta_i}\Big)}^{= \left|\frac{\partial \psi}{\partial \theta_i}\right\rangle\langle \psi| + |\psi\rangle\left\langle \frac{\partial \psi}{\partial \theta_i}\right|}{}^{\dagger} (H\rho - \rho H)\Big]$$

$$- \mathrm{Tr}\Big[(\rho H - H\rho)\Big(\frac{\partial \rho}{\partial \theta_i}\Big)\Big] =$$

$$= \left\langle \psi \middle| \frac{\partial \psi}{\partial \theta_i} \right\rangle \langle H \rangle + \left\langle \frac{\partial \psi}{\theta_i} \middle| H \middle| \psi \right\rangle$$

$$- \left\langle \psi \middle| H \middle| \frac{\partial \psi}{\partial \theta_i} \right\rangle - \left\langle \frac{\partial \psi}{\partial \theta_i} \middle| \psi \right\rangle \langle H \rangle$$

$$+ \Big(\left\langle \psi \middle| \frac{\partial \psi}{\partial \theta_i} \right\rangle \langle H \rangle + \left\langle \frac{\partial \psi}{\partial \theta_i} \middle| H \middle| \psi \right\rangle$$

$$- \left\langle \psi \middle| H \middle| \frac{\partial \psi}{\partial \theta_i} \right\rangle - \left\langle \frac{\partial \psi}{\partial \theta_i} \middle| \psi \right\rangle \langle H \rangle \Big) =$$

$$= 2 \langle H \rangle \Big[\left\langle \psi \middle| \frac{\partial \psi}{\partial \theta_i} \right\rangle - \underbrace{\left\langle \frac{\partial \psi}{\partial \theta_i} \middle| \psi \right\rangle}_{= -\left\langle \psi \middle| \frac{\partial \psi}{\partial \theta_i} \right\rangle} \Big]$$

$$+ 2 \Big[\left\langle \frac{\partial \psi}{\partial \theta_i} \middle| H \middle| \psi \right\rangle - \left\langle \psi \middle| H \middle| \frac{\partial \psi}{\partial \theta_i} \right\rangle \Big] =$$

$$= 4 \langle H \rangle \underbrace{\left\langle \psi \middle| \frac{\partial \psi}{\partial \theta_i} \right\rangle}_{= -i A_i} + 2 i \, \mathrm{Im}\Big[\left\langle \frac{\partial \psi}{\partial \theta_i} \middle| H \middle| \psi \right\rangle \Big]$$

3 irrelevant for EOM